Identification-robust inference for the LATE with high-dimensional covariates

Yukun Ma
Vanderbilt University

September 24, 2023

Motivation

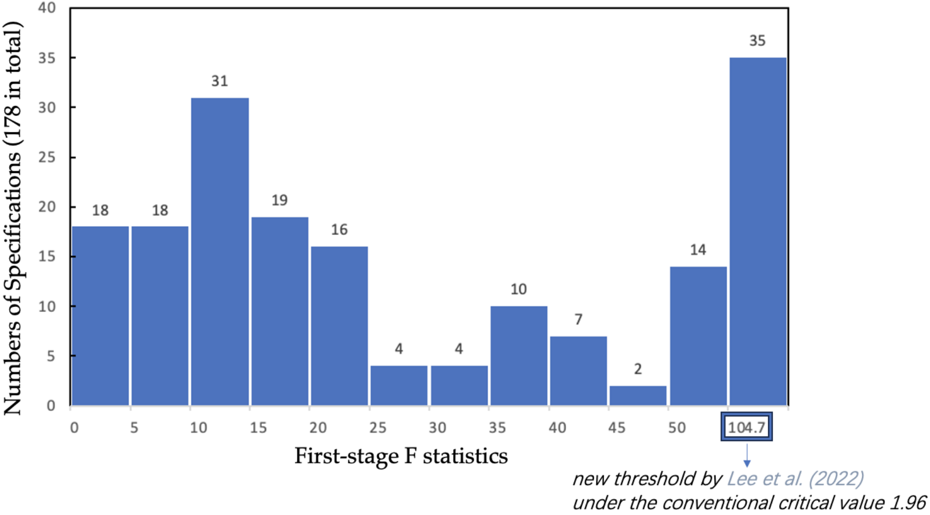


Figure: *American Economic Review* 2018-2022 ▸ heterscadesticity

Abstract

- For the local average treatment effect (LATE) with high-dimensional covariates, I develop a novel inference method, high-dimensional quasi-likelihood ratio (QLR) test, irrespective of identification strength.
    - high-dimensional controls: number of controls $\geq$ sample size.
    - weak identification: the share of compliers is small.
- The proposed method is robust to both weak identification and high dimensionality.
- The proposed test has uniformly correct asymptotic size.

Table of Contents

LATE

- the effect of a treatment for subjects who comply with the experimental treatment assigned to their sample group (compliers).
- Assume we have $\boldsymbol{N}$ observations
  - $\boldsymbol{Y_i}$: outcome of interest for unit $\boldsymbol{i}$.
  - $\boldsymbol{D_i} \in \{\mathbf{0}, \mathbf{1}\}$: receipt of treatment.
  - $\boldsymbol{Z_i} \in \{\mathbf{0}, \mathbf{1}\}$: offer of the treatment.
  - $\boldsymbol{X_i}$: $\boldsymbol{p}$-dimensional controls
    (e.g. high-dimensional covariates $\boldsymbol{p} \gg \boldsymbol{N}$).
- Imbens and Angrist (1994) propose

$$\boldsymbol{\theta} = \frac{\mathbf{E}_{P}[Y|Z=1] - \mathbf{E}_{P}[Y|Z=0]}{\mathbf{E}_{P}[D|Z=1] - \mathbf{E}_{P}[D|Z=0]} = \frac{ITT}{ITT_D} := \frac{\delta}{\pi}.$$

- Weak identification in LATE: $\boldsymbol{\pi} \to \mathbf{0}$

Weak identification

- When instruments $\boldsymbol{Z}$ are weakly correlated with endogenous regressors $\boldsymbol{D}$, conventional methods for IV estimation and inference become unreliable.

$$\boldsymbol{\theta} = \frac{\boldsymbol{\delta}}{\boldsymbol{\pi}},$$

normal approximation of $\widehat{\boldsymbol{\theta}}$ can be derived using delta method by linearized $\widehat{\boldsymbol{\theta}}$ in $(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\pi}})$. However, $\widehat{\boldsymbol{\theta}}$ is highly nonlinear in $\widehat{\boldsymbol{\pi}}$ when $\widehat{\boldsymbol{\pi}}$ is close to zero.

- Solution: test inversion.
Given $\boldsymbol{H_0} : \boldsymbol{\theta} = \boldsymbol{\theta_0}$, we have $\boldsymbol{\delta} - \boldsymbol{\theta_0}\boldsymbol{\pi} = \mathbf{0}$. Then the AR statistic

$$AR(\theta) = (\delta - \theta\pi)'\Omega(\theta)^{-1}(\delta - \theta\pi)$$

follows a χ^2 distribution under $\boldsymbol{H_0}$.

- A large number of literature in econometrics has developed methods for making inference with weak instruments,
  - Stock and Wright (2000) $\Rightarrow$ S test.
  - Kleibergen (2005) $\Rightarrow$ K test.
  - Andrews and Mikusheva (2016) $\Rightarrow$ conditional QLR test and pQLR test.
- None of them considers the model with high-dimensional covariates.

Weak identification

- When instruments $\boldsymbol{Z}$ are weakly correlated with endogenous regressors $\boldsymbol{D}$, conventional methods for IV estimation and inference become unreliable.

$$\boldsymbol{\theta} = \frac{\boldsymbol{\delta}}{\boldsymbol{\pi}},$$

normal approximation of $\widehat{\boldsymbol{\theta}}$ can be derived using delta method by linearized $\widehat{\boldsymbol{\theta}}$ in $(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\pi}})$. However, $\widehat{\boldsymbol{\theta}}$ is highly nonlinear in $\widehat{\boldsymbol{\pi}}$ when $\widehat{\boldsymbol{\pi}}$ is close to zero.

- Solution: test inversion.
Given $\boldsymbol{H_0 : \theta = \theta_0}$, we have $\boldsymbol{\delta - \theta_0 \pi = 0}$. Then the AR statistic

$$\boldsymbol{AR(\theta) = (\delta - \theta\pi)'\Omega(\theta)^{-1}(\delta - \theta\pi)}$$

follows a $\boldsymbol{\chi^2}$ distribution under $\boldsymbol{H_0}$.

- A large number of literature in econometrics has developed methods for making inference with weak instruments,
    - Stock and Wright (2000) $\Rightarrow$ S test.
    - Kleibergen (2005) $\Rightarrow$ K test.
    - Andrews and Mikusheva (2016) $\Rightarrow$ conditional QLR test and pQLR test.
- None of them considers the model with high-dimensional covariates.

Relations to the Literature: Weak Identification

Given the moment restriction $\boldsymbol{E[\phi(X;\theta_0)] = 0}$ and $\boldsymbol{g_N(\cdot) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \phi(X_i;\cdot)}$, researchers are interested in testing $\boldsymbol{H_0 : \theta = \theta_0}$.

- Stock and Wright (2000) pioneer the concepts of weakly identified GMM. They propose the S test , which is based on $\boldsymbol{S = g_N(\theta_0)'\widehat{\Sigma}(\theta_0)^{-1} g_N(\theta_0) \xrightarrow{H_0} \chi^2}$.
- Kleibergen (2005) proposes the K test that depends on the data through $\boldsymbol{g_N(\theta_0)}$ and $\boldsymbol{\frac{d}{d\theta} g_N(\theta_0)}$.
  - They focus on the processes local to the point $\boldsymbol{\theta_0}$ $\Rightarrow$ deficient power in weakly identified scenarios.
- Conditional test that based on the distribution of nonpivital statistics:
  - Moreira (2003) proposes the conditional likelihood ratio test for weakly identified linear IV models.
  - Andrews and Mikusheva (2016) develop conditional QLR test to test whether $\boldsymbol{\theta_0}$ satisfied the moment condition, without any assumption about point identification or identification strength.

Relations to the Literature: ML methods

- Belloni, Chernozhukov, and Kato (2015) advanced a Neyman orthogonal score for a Z-estimation framework in the presence of high-dimensional nuisance parameters.

- Chernozhukov et al. (2013,2016,2017) establish the CLT for high-dimensional models using the Gaussian approximation approach.

- Incorporating ML methods into the LATE framework:
  - Chernozhukov et al. (2018) introduce the double/debiased machine learning (DML) method, a combination of the Neyman orthogonality condition and cross-fitting method.
  - Belloni, Chernozhukov, Fernandez-Val, and Hansen (2017) present an efficient estimator and confidence bands for the LATE with nonparametric/high-dimensional components.

Contributions

- Weak identification in an IV context:
  - S statistic by Stock and Wright (2000), K statistic by Kleibergen (2005), Conditional test by Moreira (2003,2009), Andrews and Mikusheva (2016).
  - An important complement to existing literature on identification-robust: $\boldsymbol{p \gg N}$.
- ML based econometric methods:
  - Belloni, Chernozhukov, and Kato (2015), Chernozhukov et al. (2013,2016,2017).
  - An important complement to existing ML literature: weak identification.

Table of Contents

Setup

- Model the random vector $\boldsymbol{W} = (\boldsymbol{Y}, \boldsymbol{D}, \boldsymbol{Z}, \boldsymbol{X}')'$ as follows,

$$\begin{aligned} D &= m_0(Z, X) + v, & \mathbf{E}_P[v|Z, X] &= 0 & &\textit{First stage} \\ Y &= g_0(Z, X) + u, & \mathbf{E}_P[u|Z, X] &= 0 & &\textit{Reduced form} \\ Z &= p_0(X) + e, & \mathbf{E}_P[e|X] &= 0 & &\textit{Propensity score} \end{aligned}$$

- The LATE framework proposed by Tan (2006) is given by

$$\theta_0 = \frac{\mathbf{E}_P[g(1,X) - g(0,X) + \frac{Z}{p(X)}(Y - g(1,X)) - \frac{1-Z}{1-p(X)}(Y - g(0,X))]}{\mathbf{E}_P[m(1,X) - m(0,X) + \frac{Z}{p(X)}(D - m(1,X)) - \frac{1-Z}{1-p(X)}(D - m(0,X))]} := \frac{\mathbf{E}_P[a]}{\mathbf{E}_P[b]}$$

Anderson-Rubin-type Score

- Consider a score for LATE

$$\boldsymbol{\psi(W;\theta,\eta) = \overbrace{g(1,X) - g(0,X) + \frac{Z(Y-g(1,X))}{p(X)} - \frac{(1-Z)(Y-g(0,X))}{1-p(X)}}^{a}}$$
$$\boldsymbol{- \theta \times \underbrace{\left(m(1,X) - m(0,X) + \frac{Z(D-m(1,X))}{p(X)} - \frac{(1-Z)(D-m(0,X))}{1-p(X)}\right)}_{b},}$$

with

- target parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}$ is the LATE.
- nuisance parameter $\boldsymbol{\eta} = (\boldsymbol{g}, \boldsymbol{m}, \boldsymbol{p}) \in \boldsymbol{T}$ for a convex[1] set $\boldsymbol{T}$.

[1] To ensure that $\boldsymbol{\psi(W;\theta_0,\eta_0 + r(\eta - \eta_0))}$ is well defined for all $\boldsymbol{r} \in [\boldsymbol{0}, \boldsymbol{1}]$.

Nuisance Parameters

- Specify the nuisance parameters $\boldsymbol{\eta = (g, m, p)}$ as follows,

$$\begin{aligned}
&\boldsymbol{g(Z,X) = \mathbf{E}_P[Y|Z,X] = Z\beta_{21} + X'\beta_{22}} && \textit{Reduced form}\\
&\boldsymbol{m(Z,X) = \mathbf{E}_P[D|Z,X] = \Lambda(Z\beta_{11} + X'\beta_{12})} && \textit{First stage}\\
&\boldsymbol{p(X) = \mathbf{E}_P[Z|X] = \Lambda(X'\gamma)} && \textit{Propensity score}
\end{aligned}$$

  - The logistic CDF $\boldsymbol{\Lambda(t) = \frac{\exp(t)}{1+\exp(t)}}$ for all $\boldsymbol{t} \in \mathbb{R}$
  - The nuisance parameters $\boldsymbol{\eta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \gamma)}$.

Properties of the AR-type Score

- Moment condition:

$$\mathbf{E}_P[\underbrace{\psi(W_i;\theta_0,\eta_0)}_{a-\theta_0\times b}] = 0.$$

- Neyman orthogonality condition:
  - Path-wise (or Gateaux) derivative map D_r

$$D_r[\eta-\eta_0] := \partial_r\{\mathbf{E}_P[\psi(W;\theta_0,\eta_0+r(\eta-\eta_0))]\} \text{ for } \eta \in T.$$

  - Notation of $D_r[\eta-\eta_0]$ evaluated at $r=0$:

$$\partial_\eta \mathbf{E}_P[\psi(W;\theta_0,\eta_0)][\eta-\eta_0] := D_0[\eta-\eta_0].$$

  - The Neyman orthogonality condition holds at (θ_0,η_0) if

$$\partial_\eta \mathbf{E}_P \psi(W;\theta_0,\eta_0)[\eta-\eta_0] = 0$$

    holds for all $\eta \in \mathcal{T}_N$ for a nuisance realization set $\mathcal{T}_N \subset T$.
- The score function ψ is an AR-type Neyman orthogonal score.

High-dimensional QLR Test

Step 1: Randomly split the sample $\{1, \cdots, N\}$ into K folds $\{I_1, \cdots, I_K\}$.

Step 2: For each $k \in \{1, \cdots, K\}$, obtain $\widehat{\eta}_k$ by using only the subsample of those observations with indices $i \in \{1, \cdots, N\} \setminus I_k$: ▸ penalty parameter

(2.1) run ML (e.g., lasso) OLS regression to estimate $(\widehat{\beta}_{21}, \widehat{\beta}_{22})$,

$$(\widehat{\beta}_{21,k}, \widehat{\beta}_{22,k}) \in \arg \min_{\beta_{21},\beta_{22}} \mathbb{E}_{I_k^c}[(Y_i - Z_i\beta_{21} - X_i'\beta_{22})^2] + \frac{\lambda_3}{|I_k^c|}\|(\beta_{21}, \beta_{22})\|_1.$$

(2.2) run ML (e.g., lasso) logistic regression to estimate $(\widehat{\beta}_{11,k}, \widehat{\beta}_{12,k})$,
$(\widehat{\beta}_{11,k}, \widehat{\beta}_{12,k}) \in \arg\min_{\beta_{11},\beta_{12}} \mathbb{E}_{I_k^c}[L_1(W_i; \beta_{11}, \beta_{12})] + \frac{\lambda_1}{|I_k^c|}\|(\beta_{11}, \beta_{12})\|_1,$

(2.3) run ML (e.g., lasso) logistic regression to estimate $\widehat{\gamma}_k$,

$$\widehat{\gamma}_k \in \arg\min_{\gamma} \mathbb{E}_{I_k^c}[L_2(W_i; \gamma)] + \frac{\lambda_2}{|I_k^c|}\|\gamma\|_1,$$

- $L_1(W_i; \beta_{11}, \beta_{12}) = D_i(Z_i\beta_{11} + X_i'\beta_{12}) - \log(1 + \exp(Z_i\beta_{11} + X_i'\beta_{12}))$,
- $L_2(W_i; \gamma) = Z_i X_i'\gamma - \log(1 + \exp(X_i'\gamma))$.

High-dimensional QLR Test, Continued

Step 3: Compute $\widehat{q}_N(\theta)$ and $\widehat{\Omega}(\theta_1, \theta_2)$ for later use,

$$
\begin{aligned}
\widehat{q}_N(\theta) &= \frac{1}{\sqrt{N}} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \theta, \widehat{\eta}_k), \\
\widehat{\Omega}(\theta_1, \theta_2) &= \frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \theta_1, \widehat{\eta}_k) \psi(W_i; \theta_2, \widehat{\eta}_k) \\
&\quad - \frac{1}{N^2} \sum_{k=1}^{K} \sum_{k'=1}^{K} \sum_{i \in I_k, i' \in I_{k'}} \psi(W_i; \theta_1, \widehat{\eta}_k) \psi(W_{i'}; \theta_2, \widehat{\eta}_{k'}).
\end{aligned}
$$

An illustration of K=2-fold cross-fitting.

I_1 Score I_2 Nuisance

$\sum_{i \in I_1} \psi(W_i; \theta, \widehat{\eta}_1)$

I_1 Nuisance I_2 Score

$\sum_{i \in I_2} \psi(W_i; \theta, \widehat{\eta}_2)$

High-dimensional QLR Test, Continued

Step 4: Take independent draws $\xi \sim N(0, \widehat{\Omega}(\theta_0, \theta_0))$ and calculate $R = R(\xi, h_N, \widehat{\Omega})$, where null hypothesis H_0

$$R(\xi, h_N, \widehat{\Omega}) = \xi^2 \widehat{\Omega}(\theta_0, \theta_0)^{-1} - \inf_{\theta} (V(\theta)\xi + h_N)^2 \widehat{\Omega}(\theta, \theta)^{-1},$$

- $V(\theta) = \widehat{\Omega}(\theta, \theta_0)\widehat{\Omega}(\theta_0, \theta_0)^{-1}$
- $h_N(\theta) = \widehat{q}_N(\theta) - \widehat{\Omega}(\theta, \theta_0)\widehat{\Omega}(\theta_0, \theta_0)^{-1}\widehat{q}_N(\theta_0)$.

Step 5: Calculate the conditional critical value $c_\alpha(\tilde{h})$ as

$$c_\alpha(\tilde{h}) = \min\{c : P(R(\xi, h_N, \widehat{\Omega}) > c) \leq \alpha\}.$$

Step 6: Reject the null hypothesis $H_0 : S_N \in \mathcal{S}_0$ when $R(\xi, h_N, \widehat{\Omega})$ exceeds the $(1 - \alpha)$ quantiles $c_\alpha(h_N)$ and report the $(1 - \alpha)$ confidence interval $CI_\alpha = \{\theta : R(\xi, h_N, \widehat{\Omega}) \leq c_\alpha(h_N)\}$.

Table of Contents

Notations

- Let $c > 0$, $c_0 \geq 0$, $c_1 \geq 0$, $C_1 > 0$ be finite constants, and $a_N = p \vee N$.
- Let $\{\Delta_N\}_{N\geq 1}$, $\{\delta_N\}_{N\geq 1}$(estimation errors) be sequences of positive constants that converges to zero such that $\delta_N \geq N^{-1/2}$.
- We use $a \lesssim b$ to denote $a \leq cb$ for some $c > 0$ that does not depends on N.
- Let $\|\delta\|_0$ represent the number of non-zero components of δ.
- Let $\mathcal{P}_N$ be the probability law of $\{W_i\}_{i=1}^N$.
- Let $\mathcal{P}_0$ be family of distribution consistent with the null.
- The sequence $\{M_N\}_{N\geq 1}$ be a set of positive constants such that $M_N \geq (\mathbf{E}_P[(Z_i \vee \|X_i\|_\infty)^{2q}])^{1/2q}$.

Assumption: Regularity Conditions for the LATE

For $P \in \mathcal{P}_N$, the following conditions hold.

(i) The following equations are satisfied with a binary D and Z.

$$\left.\begin{array}{ll} D = m_0(Z, X) + v, & \mathrm{E}_P[v|Z, X] = 0 \\ Y = g_0(Z, X) + u, & \mathrm{E}_P[u|Z, X] = 0 \end{array}\right\} \rightarrow \begin{array}{c}\text{Independence.} \\ (Y, D)|X \perp\!\!\!\perp Z\end{array}$$

$$Z = p_0(X) + e, \quad \mathrm{E}_P[e|X] = 0 \quad \rightarrow \text{Exclusion Restriction.}$$

(ii) For some $\varepsilon > 0$, $\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon$ almost surely.

(iii) Θ is compact.

(iv) $\mathrm{E}_P[D|Z = 1] \geq \mathrm{E}_P[D|Z = 0]$. ▸ Assumption Comparison

(v) $\|u\|_{P,2} \geq c_0$, and $\|\mathrm{E}_P[u^2|X]\|_{P,\infty} \leq c_1$.

Assumption: Nuisance Parameter Estimators

- Sparse eigenvalue conditions: with probability $1 - o(1)$, for some $l_N \to \infty$ slow enough, we have

$$1 \lesssim \phi_{\min}(l_N s_N) \leq \phi_{\max}(l_N s_N) \lesssim 1.$$

▸ spare eigenvalue

- Sparsity: $\|\beta_{12}^0\|_0 + \|\beta_{22}^0\|_0 + \|\gamma^0\|_0 \leq s_N$.
- Parameters: $\|\beta_{12}^0\| + \|\beta_{22}^0\| + \|\gamma^0\| \leq C_1$.
- Covariates: for $q > 4$,
  - $\inf_{\|\xi\|=1} \mathrm{E}_P[((Z_i, X_i')\xi)^2] \geq c$.
  - $\sup_{\|\xi\|=1} \mathrm{E}_P[((Z_i, X_i')\xi)^2] \leq C_1$.
  - $N^{-1/2+2/q} M_N^2 s_N \log^2 a_N \leq \Delta_N$.

Main result

The empirical process

$$\mathbb{G}_N(\cdot) = \underbrace{\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \big(\psi(W_i; \cdot, \eta_0)}_{q_N(\theta)} - \mathbf{E}_P[\psi(W; \cdot, \eta_0)]\big).$$

Propose an estimator of $\mathbb{G}_N(\cdot)$ as

$$\widehat{\mathbb{G}}_N(\theta) = \underbrace{\sqrt{N}\Big(\frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \theta, \widehat{\eta}_k)}_{\widehat{q}_N(\theta)} - \mathbf{E}_P\left[\psi(W_i; \theta, \widehat{\eta}_k)\right]\Big).$$

Theorem

Suppose that the above assumptions are satisfied. Under the null, we have

$$\widehat{\mathbb{G}}_N(\theta) = \mathbb{G}_N(\theta) + O_P(N^{-1}).$$

The process $\widehat{\mathbb{G}}_N(\cdot)$ weakly converges to a centered Gaussian process $\mathbb{G}(\cdot)$ over $P \in \mathcal{P}_0$ as $N \to \infty$ with covariance function $\Omega(\theta_1, \theta_2) =$ $\mathbf{E}_P[(\psi(W; \theta_1, \eta_0) - \mathbf{E}_P[\psi(W; \theta_1, \eta_0)]) (\psi(W; \theta_2, \eta_0) - \mathbf{E}_P[\psi(W; \theta_2, \eta_0)])]$.

Theorem: Variance Estimation

Under the same set of assumptions as above, the variance $\boldsymbol{\Omega}(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ can be consistently estimated uniformly over $\boldsymbol{P} \in \boldsymbol{\mathcal{P}_0}$ by

$$\begin{aligned}\widehat{\boldsymbol{\Omega}}(\theta_1, \theta_2) &= \frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \psi(W_i; \theta_1, \widehat{\eta}_k) \psi(W_i; \theta_2, \widehat{\eta}_k) \\ &\quad - \frac{1}{N^2} \sum_{k,k'=1}^{K} \sum_{i \in I_k, i' \in I_{k'}} \psi(W_i; \theta_1, \widehat{\eta}_k) \psi(W_{i'}; \theta_2, \widehat{\eta}_{k'})\end{aligned}$$

and $\widehat{\boldsymbol{\Omega}}(\theta_1, \theta_2) = \boldsymbol{\Omega}(\theta_1, \theta_2) + O_P(\rho_N)$ with $\rho_N \lesssim \delta_N$.

How It Works

- The equivalence between bounded Lipschitz convergence and weak convergence of stochastic processes. See Section 1.12 in van der Vaart and Wellner (1996).

- Weak convergence:

Step 1 the convergence of the finite dimensional distribution of $\widehat{\mathbb{G}}_{\boldsymbol{N}}(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta_I}$.

Step 2 the asymptotic equicontinuity of $\widehat{\mathbb{G}}_{\boldsymbol{N}}(\boldsymbol{\theta})$ over $\boldsymbol{P} \in \boldsymbol{\mathcal{P}_0}$,

$$\lim_{N\to\infty} \lim_{|\theta_1-\theta_2|\to 0} P\left(|\widehat{\mathbb{G}}_N(\theta_1) - \widehat{\mathbb{G}}_N(\theta_2)| > \epsilon_1\right) = 0$$

for any $\boldsymbol{\varepsilon_1 > 0}$.

Step 3 the boundedness of $\boldsymbol{\Theta_I}$.

Theorem: Size Control

Under the same set of assumptions above, the test that rejects the null hypothesis $\boldsymbol{H_0 : S_N \in \mathcal{S}_0}$ when $\boldsymbol{R(q_N(\theta_0), h_N, \Omega)}$ exceeds the $\boldsymbol{(1-\alpha)}$ quantile $\boldsymbol{c_\alpha(h_N)}$ of its conditional distribution given $\boldsymbol{h_N(\cdot)}$ has uniformly correct asymptotic size. Under the null, we have

$$\lim_{N\to\infty} \sup_{P\in\mathcal{P}_0} P(R(\widehat{q}_N(\theta_0), h_N, \widehat{\Omega}) > c_\alpha(h_N)) = \alpha.$$

Table of Contents

Simulation Setup

- Primitive random vector $\boldsymbol{X}_i'$ is constructed by

$$\boldsymbol{X}_i \sim \boldsymbol{N}\left(\boldsymbol{0}, \begin{pmatrix} U^0 & U^1 & \cdots & U^{\dim(X)-2} & U^{\dim(X)-1} \\ U^1 & U^0 & \cdots & U^{\dim(X)-3} & U^{\dim(X)-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ U^{\dim(X)-2} & U^{\dim(X)-3} & \cdots & U^0 & U^1 \\ U^{\dim(X)-1} & U^{\dim(X)-2} & \cdots & U^1 & U^0 \end{pmatrix}\right)$$

  with $\boldsymbol{U} = \mathbf{0.5}$.

- Consider $\boldsymbol{N} = \mathbf{500}$, $\mathbf{dim}(\boldsymbol{X}) = \mathbf{5}, \underbrace{\mathbf{100}, \mathbf{300}, \text{and } \mathbf{500}.}_{\text{high-dimensional LATE}}$

- Define the compliance class $\boldsymbol{Q_i} := \begin{cases} \mathbf{0} & \text{never-taker} \\ \mathbf{1} & \text{always-taker} \\ \mathbf{2} & \text{compliers} \end{cases}$

- The compliance score denoted as $\boldsymbol{\delta(x)}$ is constructed as $\boldsymbol{\delta(x)} = \boldsymbol{P}[\boldsymbol{Q_i} = \mathbf{2} | \boldsymbol{X} = \boldsymbol{x}] = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$.

Simulation Setup, Continued

- The probability of being a never-taker or always-taker is $(1-\delta(x))/2$.
- The parameter values (β_0, β_1) are set such that

$$P(Q_i = 2) = \begin{cases} 0.1 & \text{weakly identified case} \\ 0.5 & \text{strongly identified case} \end{cases}$$

- Generate random variables (Z_i, D_i, Y_i) as follows
  - $Z_i = \frac{\exp(\gamma_0+\gamma_1 x)}{1+\exp(\gamma_0+\gamma_1 x)} + v_i \overset{s.t.}{\Longrightarrow} P(Z_i = 1) = 0.5$.
  - $D_i = Z_i * \mathbb{1}\{Q_i = 2\} + Q_i * \mathbb{1}\{Q_i \neq 2\}$.
  - $Y_i = D_i\theta_0 + X_i + \varepsilon_i \quad \Longrightarrow \theta_0 = 1.0$.
  - v_i, ε_i are independently generated according to $v_i, \varepsilon_i \sim N(0,1)$.

Results

I compare the proposed method HD-QLR (this paper) with

- the conditional QLR test (AM16)[2] : robust against weak identification but not against high dimensionality.
- ML methods (CCDDHNR18[3] and BCFH17[4]): robust against high dimensionality but not against weak identification.

▸ Approaches Comparison

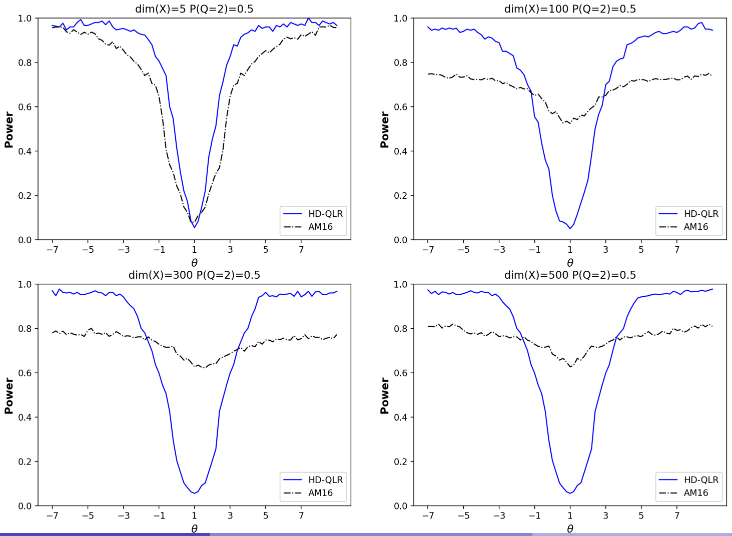
[2] Andrews and Mikusheva (2016).

[3] Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018).

[4] Belloni, Chernozhukov, Fernandez, and Hansen (2017).

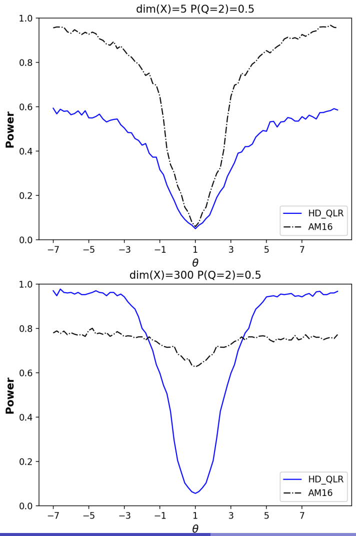
Comparisons: strong identification ▸ Power Curve

- Power curve of nominal 5%: AM16 with HD-QLR (this paper)



Comparisons: weak identification

- Power curve of nominal 5%: AM16 with HD-QLR (this paper)



Results

I compare the proposed method HD-QLR (this paper) with

- the conditional QLR test (AM16)[2] : robust against weak identification but not against high dimensionality.
- ML methods (CCDDHNR18[3] and BCFH17[4]): robust against high dimensionality but not against weak identification.

Approaches Comparison

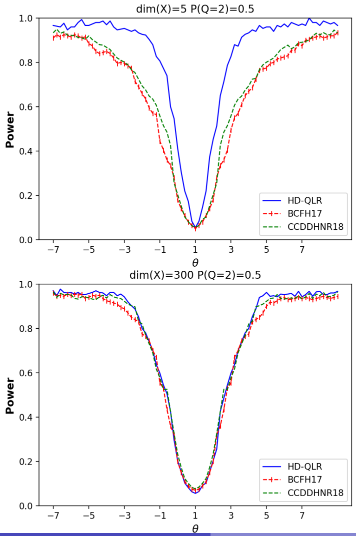
[2] Andrews and Mikusheva (2016).

[3] Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018).

[4] Belloni, Chernozhukov, Fernandez, and Hansen (2017).

Comparisons: strong identification

- Power curve: CCDDHNR18, BCFH17 with HD-QLR (this paper)



Comparisons: weak identification

- Power curve: CCDDHNR18, BCFH17 with HD-QLR (this paper)

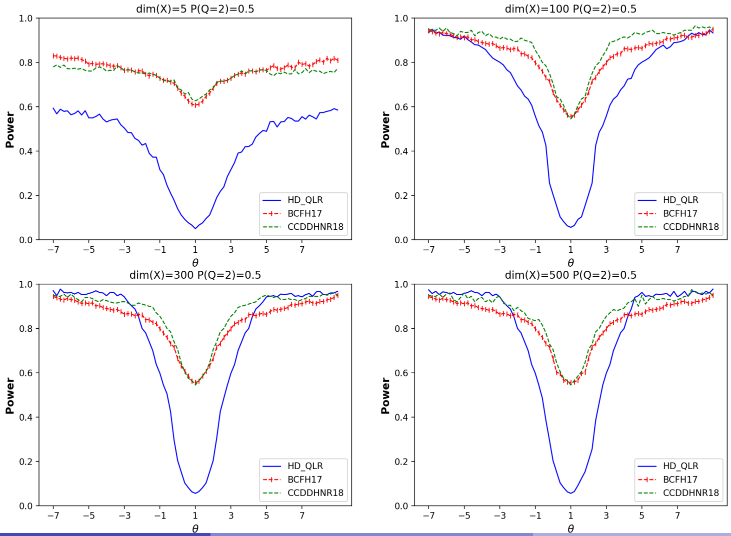


Table of Contents

Example One: Hornung (2015)
"Railroads and growth in Prussia"

Data: highly detailed city-level data from the historical German state of Prussia.

- Y_{it} : urban population growth rate during time period t.
- D_i : whether the city i was connected to the railroad in 1848.
- Z_i : whether the city i was located within a straight-line corridor between two important cities.
- X_i : whether the city had street access, whether the city had waterway access, military population, age composition, school enrollment rate, etc.

Results

$\boldsymbol{Y_{it}}$: population growth rate	periods						
	49-52	52-55	55-58	58-61	61-64	64-67	67-71
Panel A: AM16							
LATE	0.010	0.020	0.063	0.030	0.037	0.056	0.044
CI	[-0.017, 0.05]	[0.004, 0.039]	[0.030, 0.063]	[0.011, 0.050]	[0.019, 0.050]	[0.012, 0.420]	[0.018, 0.155]
length of CI	0.067	0.035	0.033	0.039	0.031	0.408	0.137
Panel B: CCDDHNR18							
LATE	0.012	0.011	0.007	0.000	0.020	0.012	0.011
CI	[-0.019, 0.039]	[-0.014, 0.035]	[-0.009, 0.044]	[-0.016, 0.030]	[-0.019, 0.052]	[-0.014, 0.039]	[-0.016, 0.036]
length of CI	0.058	0.048	0.053	0.046	0.070	0.052	0.052
Panel C: BCFH17							
LATE	0.009	0.009	0.012	0.006	0.015	0.012	0.013
CI	[-0.009, 0.026]	[-0.006, 0.023]	[-0.007, 0.031]	[-0.008, 0.020]	[-0.009, 0.040]	[-0.018, 0.041]	[-0.008, 0.034]
length of CI	0.035	0.029	0.038	0.028	0.049	0.059	0.042
Panel D: HD-QLR (this paper)							
LATE	0.010	0.011	0.014	0.004	0.018	0.014	0.011
CI	[0.000, 0.021]	[0.002, 0.018]	[0.003, 0.027]	[-0.001, 0.016]	[0.003, 0.029]	[-0.004, 0.032]	[-0.002, 0.023]
length of CI	0.021	0.016	0.024	0.017	0.026	0.033	0.024
Size N	929	924	914	926	924	919	919
dim(X)	212	212	212	212	212	212	212

Results, Continued

$\boldsymbol{Y_{it}}$: population growth rate	periods						
	49-52	52-55	55-58	58-61	61-64	64-67	67-71
Panel A: AM16							
LATE	0.010	0.020	0.063	0.030	0.037	0.056	0.044
CI	[-0.017, 0.05]	[0.004, 0.039]	[0.030, 0.063]	[0.011, 0.050]	[0.019, 0.050]	[0.012, 0.420]	[0.018, 0.155]
length of CI	0.067	0.035	0.033	0.039	0.031	0.408	0.137
Panel B: CCDDHNR18							
LATE	0.012	0.011	0.007	0.000	0.020	0.012	0.011
CI	[-0.019, 0.039]	[-0.014, 0.035]	[-0.009, 0.044]	[-0.016, 0.030]	[-0.019, 0.052]	[-0.014, 0.039]	[-0.016, 0.036]
length of CI	0.058	0.048	0.053	0.046	0.070	0.052	0.052
Panel C: BCFH17							
LATE	0.009	0.009	0.012	0.006	0.015	0.012	0.013
CI	[-0.009, 0.026]	[-0.006, 0.023]	[-0.007, 0.031]	[-0.008, 0.020]	[-0.009, 0.040]	[-0.018, 0.041]	[-0.008, 0.034]
length of CI	0.035	0.029	0.038	0.028	0.049	0.059	0.042
Panel D: HD-QLR (this paper)							
LATE	0.010	0.011	0.014	0.004	0.018	0.014	0.011
CI	[0.000, 0.021]	[0.002, 0.018]	[0.003, 0.027]	[-0.001, 0.016]	[0.003, 0.029]	[-0.004, 0.032]	[-0.002, 0.023]
length of CI	0.021	0.016	0.024	0.017	0.026	0.033	0.024
Size N	929	924	914	926	924	919	919
dim(X)	212	212	212	212	212	212	212

Results, Continued

Y_{it}: population growth rate	49-52	52-55	55-58	58-61	61-64	64-67	67-71
	periods						
Panel A: AM16							
LATE	0.010	0.020	0.063	0.030	0.037	0.056	0.044
CI	[-0.017, 0.05]	[0.004, 0.039]	[0.030, 0.063]	[0.011, 0.050]	[0.019, 0.050]	[0.012, 0.420]	[0.018, 0.155]
length of CI	0.067	0.035	0.033	0.039	0.031	0.408	0.137
Panel B: CCDDHNR18							
LATE	0.012	0.011	0.007	0.000	0.020	0.012	0.011
CI	[-0.019, 0.039]	[-0.014, 0.035]	[-0.009, 0.044]	[-0.016, 0.030]	[-0.019, 0.052]	[-0.014, 0.039]	[-0.016, 0.036]
length of CI	0.058	0.048	0.053	0.046	0.070	0.052	0.052
Panel C: BCFH17							
LATE	0.009	0.009	0.012	0.006	0.015	0.012	0.013
CI	[-0.009, 0.026]	[-0.006, 0.023]	[-0.007, 0.031]	[-0.008, 0.020]	[-0.009, 0.040]	[-0.018, 0.041]	[-0.008, 0.034]
length of CI	0.035	0.029	0.038	0.028	0.049	0.059	0.042
Panel D: HD-QLR (this paper)							
LATE	0.010	0.011	0.014	0.004	0.018	0.014	0.011
CI	[0.000, 0.021]	[0.002, 0.018]	[0.003, 0.027]	[-0.001, 0.016]	[0.003, 0.029]	[-0.004, 0.032]	[-0.002, 0.023]
length of CI	0.021	0.016	0.024	0.017	0.026	0.033	0.024
Size N	929	924	914	926	924	919	919
dim(X)	212	212	212	212	212	212	212

Example Two: Ambrus, Field, and Gonzalez (2020) "The Impact on Housing Prices of A Cholera Epidemic"

"In August 1854, St. James experienced a sudden outbreak of cholera when one of the 13 shallow wells that serviced the parish, the Broad Street pump, became contaminated with cholera bacteria."

- $\boldsymbol{Y_i}$ **:** the log rental price of house $\boldsymbol{i}$ in 1864.
- $\boldsymbol{D_i}$ **:** whether house $\boldsymbol{i}$ had at least one cholera death.
- $\boldsymbol{Z_i}$ **:** whether house $\boldsymbol{i}$ fell inside the contaminated areas.
- $\boldsymbol{X_i}$ **:** distance to the closest pump, distance to the fire station, distance to the urinal, sewer access, among a total of 23 variables.

Results

	AM16	CCDDHNR18	BCFH17	HD-QLR (this paper)
LATE	-1.205	-0.413	-0.357	-0.421
CI	[-2.230,	[-3.565,	[-1.291,	[-1.080,
	-0.650]	2.670]	0.576]	0.035
length of CI	1.580	6.235	1.866	1.115

Table: Displayed are LATE estimates, CIs and the length of CI for the coefficient of the cholera-related deaths using four different approaches: AM16, CCDDHNR18, BCFH17, and the proposed HD-QLR. Estimation and inference results in CCDDHNR18 and HD-QLR are based on 10 iterations of resampled cross fitting with $\boldsymbol{K = 4}$ folds for cross fitting. The number of observations $\boldsymbol{N = 467}$.

Table of Contents

Takeaways

- I develop a test statistic to make inference for the high-dimensional LATE, independent of the strength of identification.

	Low-dimensional Model	High-dimensional Model
Strong Identification	t-test	CCDDHNR18, BCFH17
Weak Identification	AR, S, K, AM16	HD-QLR (my paper)

- The test has uniformly correct asymptotic size.
- Simulation results indicate that the proposed test is robust against weak identification and high-dimensional settings, outperforming other conventional tests.
- Empirical illustrations show that conventional tests exhibit a positive bias in the length of confidence intervals and lose significance when high-dimensional covariates are taken into account.

Takeaways

- I develop a test statistic to make inference for the high-dimensional LATE, independent of the strength of identification.

	Low-dimensional Model	High-dimensional Model
Strong Identification	t-test	CCDDHNR18, BCFH17
Weak Identification	AR, S, K, AM16	HD-QLR (my paper)

- The test has uniformly correct asymptotic size.
- Simulation results indicate that the proposed test is robust against weak identification and high-dimensional settings, outperforming other conventional tests.
- Empirical illustrations show that conventional tests exhibit a positive bias in the length of confidence intervals and lose significance when high-dimensional covariates are taken into account.

Thank you!

feel free to email me any comments
yukun.ma@vanderbilt.edu

Motivation: Lee et al. (2022)

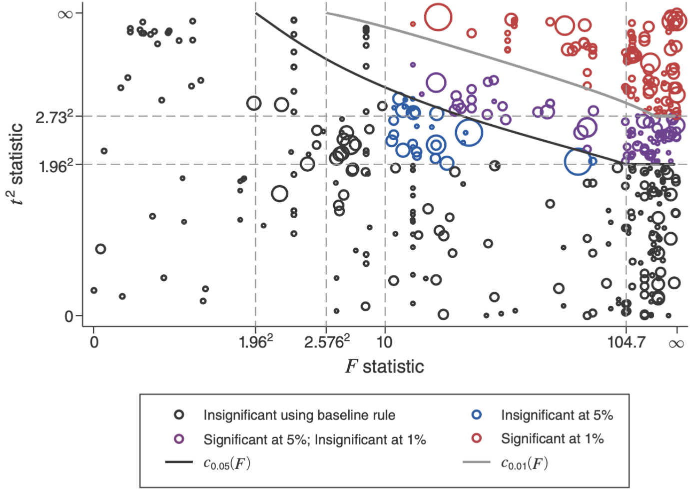


Figure: *American Economic Review* 2013-2019

back

Tuning Parameters

Lemma (Convergence rate for Lasso with logistic model)

Suppose some regularity assumptions hold. In addition, suppose that the penalty choice $\boldsymbol{\lambda_1 = K_1 \sqrt{N \log(pN)}}$ *and* $\boldsymbol{\lambda_2 = K_2 \sqrt{N \log(pN)}}$ *for* $\boldsymbol{K_1, K_2 > 0}$. *Then with probability* $\boldsymbol{1 - o(1)}$,

$$\|(\widehat{\beta}_{11}, \widehat{\beta}_{12}) - (\beta^0_{11}, \beta^0_{12})\| \vee \|\widehat{\gamma} - \gamma^0\| \lesssim \sqrt{\frac{s_N \log(pN)}{N}}.$$

Lemma (Convergence rate for Lasso with OLS)

Suppose some regularity assumptions hold. Moreover, suppose that the penalty choice $\boldsymbol{\lambda_3 = K_3 \sqrt{N \log(pN)}}$ *for* $\boldsymbol{K_3 > 0}$. *Then with probability* $\boldsymbol{1 - o(1)}$,

$$\|(\widehat{\beta}_{21}, \widehat{\beta}_{22}) - (\beta^0_{21}, \beta^0_{22})\| \lesssim \sqrt{\frac{s_N \log(pN)}{N}}.$$

back

Null Hypothesis

- Define $\boldsymbol{S_N(\cdot) = \mathbf{E}_P[N^{-1/2}\sum_{i=1}^{N}\psi(W_i;\cdot,\eta_0)]}$.

- $\boldsymbol{H_0 : \theta = \theta_0}$ or $\boldsymbol{\theta_0 \in \Theta_I}$ with the identified set $\boldsymbol{\Theta_I \subset \Theta}$.

$$\Updownarrow$$

$$\boldsymbol{S_N(\theta_0) = 0}.$$

- Let $\boldsymbol{\mathcal{S}_0}$ be the collection of function $\boldsymbol{S_N(\cdot)}$ satisfying $\boldsymbol{S_N(\theta_0) = 0}$.

$$\Updownarrow$$

$$\boldsymbol{H_0 : S_N(\cdot) \in \mathcal{S}_0}.$$

back

Spare eigenvalue

For any $\boldsymbol{T} \subset [\boldsymbol{p}+\mathbf{1}]$, $\boldsymbol{\delta} = (\boldsymbol{\delta_1}, \cdots, \boldsymbol{\delta_{p+1}})' \in \mathbb{R}^{\boldsymbol{p}+\mathbf{1}}$ with $\boldsymbol{\delta_{T,j}} = \boldsymbol{\delta_j}$ if $\boldsymbol{j} \in \boldsymbol{T}$ and $\boldsymbol{\delta_{T,j}} = \mathbf{0}$ if $\boldsymbol{j} \notin \boldsymbol{T}$. Define the minimum and maximum sparse eigenvalue by

$$\phi_{\min}(m) = \inf_{\|\delta\|_0 \le m} \frac{\|(Z_i, X_i')\delta\|_{2,N}}{\|\delta_T\|_1}$$

$$\phi_{\max}(m) = \sup_{\|\delta\|_0 \le m} \frac{\|(Z_i, X_i')\delta\|_{2,N}}{\|\delta_T\|_1}.$$

back

Identification Assumptions

- Assumption in my paper:

$$\boldsymbol{Ep[D|Z=1] \geq \mathbf{E}_P[D|Z=0].}$$

- Assumption in weak identification literature:

$$\boldsymbol{Ep[D|Z=1] - \mathbf{E}_P[D|Z=0] = \frac{C_1}{\sqrt{N}}} \quad \text{with } \boldsymbol{C_1 > 0.}$$

- Assumption in ML literature:

$$\boldsymbol{Ep[D|Z=1] - \mathbf{E}_P[D|Z=0] \geq C_2} \quad \text{with } \boldsymbol{C_2 > 0.}$$

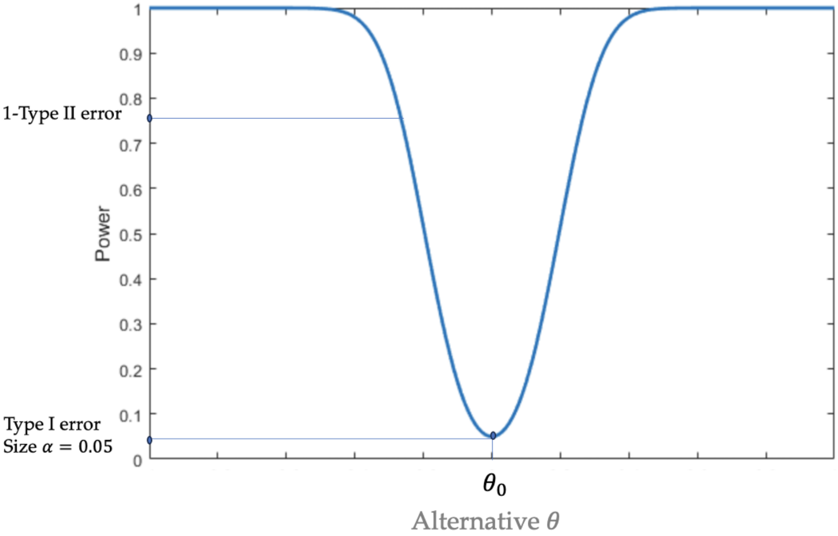
back

Comparison Across Four Approaches

- AM16
  - Neyman orthogonal score $\boldsymbol{\psi}$.
  - test statistics $\boldsymbol{R}$.
  - use traditional methods to handle $\boldsymbol{X_i}$.
- CCDDHNR18 BCFH17
  - Neyman orthogonal score $\boldsymbol{\psi}$.
  - normal t-test.
  - use ML to handle $\boldsymbol{X_i}$.
- Drawbacks
  - overfitting bias;
  - capturing noise;
  - multicollinearity;
  - cannot perform well under weakly identified scenarios.
- The proposed HD-QLR takes advantages from both methods:
  - Neyman orthogonal score $\boldsymbol{\psi}$.
  - test statistics $\boldsymbol{R}$.
  - use ML (Lasso) to handle the high-dimensional $\boldsymbol{X_i}$.

back

Power Curve



back

Ambrus, A., E. Field, and R. Gonzalez (2020): "Loss in the time of cholera: Long-run impact of a disease epidemic on the urban landscape," *American Economic Review*, 110, 475–525.

Andrews, I. and A. Mikusheva (2016): "Conditional inference with a functional nuisance parameter," *Econometrica*, 84, 1571–1612.

Belloni, A., V. Chernozhukov, I. Fernandez-Val, and C. Hansen (2017): "Program evaluation and causal inference with high-dimensional data," *Econometrica*, 85, 233–298.

Belloni, A., V. Chernozhukov, and K. Kato (2015): "Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems," *Biometrika*, 102, 77–94.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022): "When is TSLS actually late?" Tech. rep., National Bureau of Economic Research.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013): "Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors," *The Annals of Statistics*, 41, 2786–2819.

——— (2016): "Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings," *Stochastic Processes and their Applications*, 126, 3632–3651.

——— (2017): "Central limit theorems and bootstrap in high dimensions," *The Annals of Probability*, 45, 2309–2352.

Hornung, E. (2015): "Railroads and growth in Prussia," *Journal of the European Economic Association*, 13, 699–736.

Kleibergen, F. (2005): "Testing parameters in GMM without assuming that they are identified," *Econometrica*, 73, 1103–1123.

Moreira, M. J. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

Stock, J. H. and J. H. Wright (2000): "GMM with weak identification," *Econometrica*, 68, 1055–1096.

Tan, Z. (2006): "Regression and weighting methods for causal inference using instrumental variables," *Journal of the American Statistical Association*, 101, 1607–1618.

van der Vaart, A. and J. A. Wellner (1996): "WEAK CONVERGENCE AND EMPIRICAL PROCESSES," .